

# **Cambridge Recycling Participation Study Statistical Analysis**

Submitted by:  
Clear View Consulting  
164 West River Street  
Orange, MA 01364  
(978) 544-5872

## **Introduction**

This report attempts to provide a framework for understanding the statistical significance of the results of the Cambridge Recycling Participation Study. Although some material from previous reports in this study has been repeated where directly necessary for development of the argument being made, the reader's attention is directed to those previous reports for a greater understanding of the field methodologies employed. To summarize, this study encountered many complicating factors that could not be effectively isolated within the limited scope of the study. In spite of these complications, it appears reasonable to conclude that the outreach methods being tested were effective. Like all other outreach methods that have been utilized to promote recycling participation, "effective" does not mean universally effective.

## **Statement of Problem**

The primary research question of this study is whether two forms of outreach utilizing the principles of Community-Based Social Marketing (CBSM) have a statistically significant effect on increasing recycling behavior among previously non-recycling households. As is often the case, in the real world, designing a "clean" study that neatly measures the effect in question while controlling for all other variables is easier said than done. This section describes the original study design, what happened as it was implemented, the list of other factors, which must be controlled for, and the somewhat modified analysis plan that resulted.

The two CBSM techniques to be tested were (a) door-to-door, and (b) telephone appeals by members of 2 East Cambridge community groups, asking residents to recycle for all of the standard reasons, and offering them the added inducement that the groups would receive a small financial benefit if the targeted households began to recycle. A third "treatment" group would receive a special mailing, while a fourth would function as a control group. As originally planned, the study was to identify 600 non-recycling households, 150 of which would be assigned to each of the four treatment groups.

Various difficulties during the "before" monitoring presented a difficult choice between accepting a slightly smaller study group of 567 households (even that number was reached only by including 32 buildings in which there was one unidentified recycling

household), or extending the “before” monitoring phase to allow the acquisition of a better sample. Because of pressure to meet the schedule for the Outreach phase, the decision was made to accept the smaller sample, and slightly smaller treatment group sizes ranging from 140 to 143.

The next difficulty encountered was greater than expected trouble obtaining phone numbers for the households in that treatment group. Between unlisted numbers, out-of-date listings, and difficulty getting people to answer the phone, fewer than half the households in this group were even contacted during the outreach phase. (For both the phone and door-to-door treatment groups, the study design called for three attempts to reach each household.)

Another difficulty that arose during the Outreach phase (which had been anticipated, but not nearly to the degree that it occurred) was the number of households stating that they were already recycling. Some number of these undoubtedly really were recycling, and were either among the 32 unidentified recycling households mentioned above, or were households whose recycling behavior had been missed during the “before” monitoring (described as “false negatives” in the analysis to follow). Based on experience from previous studies, some may also have counted returning deposit containers as recycling behavior, and others may have reported it simply because they perceived it as the socially correct thing to say.

All of these factors resulted in both of the CBSM outreach treatment groups fragmenting into four subgroups: the “committers” (what we really wanted to study), the “already recycling”, the “refused”, and the “not reached”. As previously described in the write-up of Final Results, during the outreach and both follow-up monitoring phases a number of buildings were identified that no longer fit the criteria for the study. When these were set aside, we were left with 511 households. Luckily, only one out of 35 “committer” households was among those that had to be dropped.

The results to be analyzed are shown in the three graphs in Attachment 1. Whereas in the Final Results write-up, the directly observed percentages recycling in each phase were shown, for this report the documented “before” recycling percentages have been subtracted out to produce a NET increase in recycling for each group and subgroup for both the “3-month” and “12-month” monitoring. For each group or subgroup, the final sample size is shown as “N= \_\_\_” in the bar labels.

The first of the three graphs compares the four groups that form the key comparison of the original study design: the door-to-door and phone “committers”, the “mailing” (less those households for which the brochure was returned as undeliverable), and the control group. Upon initial inspection, it looks fairly encouraging, except for the drop-out of 3 out of 4 phone “committers” between the 3- and 12-month monitoring. It is the second and third graphs, however, that show us that there are other things going on that we will have to account for if we hope to statistically interpret these results. Each of these graphs shows three other subgroups with notably higher end recycling increases than either the control or the mailing group. In total, four of the subgroups show a higher

end result than the phone “committers” subgroup, and one subgroup looks fairly close to the end result for the door “committers”.

Fortunately, during the monitoring process, we were able to identify several other factors that might be at work. While data on these factors that seems relevant to the East Cambridge study area is limited, the next section of this report describes them and attempts to assemble them into a theoretical framework to carry this analysis forward. Once this framework is established, the actual data will be analyzed in that context.

### **Theoretical Derivation of “Null Hypothesis”**

The goal of this study is somewhat analogous to a study of a new experimental drug: can the results obtained with it be shown to be better than what would have happened without it, in a statistically significant way? In a statistical sense, the “what would have happened without it” is called the “null hypothesis”, which in this case is considerably different from “nothing” and requires further definition. The “alternative hypothesis”, which is that the CBSM outreach methods produce a significantly better outcome (in the form of a greater increase in recycling behavior) is what we seek to test once we have more fully defined the parameters of the “null hypothesis”.

During the various phases of this study, Clear View Consulting and/or Cambridge DPW staff walked up and down each of the various streets in the study area at least 20 times. During that time, in addition to the specific data that was being sought for the study, it was inevitable that other factors relevant to the outcome of the study, but not explicitly provided for in the study design, were observed. The largest of these by far was the scope and apparent effect of household turnover. Census statistics show that for Cambridge as a whole, only 38.7% of households in 2000 had lived in the same house in 1995. This translates to an average annual turnover rate of 17.3%. While statistics specific to this study area were not readily available, it seems almost certain that the study area would match, and likely exceed, the citywide average.

Household turnover is of course not distributed evenly throughout the year. While exact numbers are not available, turnover (especially in a city like Cambridge with a large student population) is concentrated around September 1 and June 1. In the course of this study, there would have been a “June 1” turnover during the later stages of the Outreach phase, a “September 1” turnover just before the beginning of the “3-month” monitoring, and another “June 1” turnover during the “12-month” monitoring. Therefore, even the citywide average turnover rate would have produced total turnover in excess of 20% in the study area.

The rate of turnover by itself does not define the impact on recycling rates. One may presume that an “average” household moving into a building has a certain propensity toward recycling, derived from experience with it in other communities or another living situation in the same community. This likelihood of recycling must surely be at or above 50%, or the overall level of recycling in the city would not be sustained.

In a sample of non-recycling households, annual turnover in the range of 17.5 to 20% coupled with a 50%+ likelihood that the new households will recycle will produce a 10%+ increase in the recycling rate observed in the sample the first year, with a gradually slowing rate of increase each year as the 50% level is approached.

A more complete listing of the factors which were identified during the study and/or from previous studies appears as page 1 of Attachment 2. Another factor to consider is the possibility that other Cambridge outreach efforts affected households in this study at some point during the nearly 15 months from the beginning of the “before” monitoring to the end of the “12-month” monitoring. There is also considerable anecdotal evidence (although CVC is not aware of any attempts to quantify it) of a “peer effect”, in which if enough other households in their building or on their street start to recycle, a household may be more inclined to join in. Also, there is a possibility that some study households, becoming aware of the ongoing study, are influenced to change their recycling behavior.

In addition to the above factors that involve changes in household behavior, it is important to consider the possibility of measurement error in this or any other field study of recycling behavior. While considerable emphasis was placed on procedures to make the assessment of recycling set-outs both thorough and consistent, in a dense urban area like East Cambridge, possibilities for error remain. One category of error is a “false positive”, in which a household is credited with recycling behavior when it is not actually recycling. A “false negative” occurs when a household’s recycling behavior is not noted. As noted in the Final Results write-up, it seems almost inevitable that some of both occurred during this study. In a study of this type, the each type of error would have an opposite effect on the overall results depending on whether it occurred in the “before” or “after” monitoring.

While all of these effects are likely to be operating in the larger neighborhood in which this study was conducted, they necessarily do not affect each individual household equally. Thus, when samples are drawn from the larger population, a different result might be obtained each time. Generally, the likelihood of various results can be defined by a mean, or average, result and a standard deviation, or measure of variation. Although we have no way of proving it conclusively, it seems reasonable that the distribution of effects that in combination we will call our null hypothesis would be close to the standard normal distribution, a bell-shaped curve. Since we have many factors operating and imperfect information about many of them, it was decided to try to “construct” the “null hypothesis” by defining its mean and standard deviation from a series of variously well- or less-well educated guesses. We will then analyze the relevant study subgroups and compare the mean and standard deviation obtained from that data to our theoretical distribution.

Page 2 of Attachment 2 illustrates the attempt to construct a theoretical distribution for the “null hypothesis”. This was done separately for the “3-month” monitoring period and the “12-month” monitoring period. As described above, the household turnover effect is the largest effect to be accounted for. Its contribution to the

mean predicted change was estimated at 11% for the 3-month period and 13.8% for the 12-month period. The possible impact of other outreach was estimated at a rather small 1% at 3 months, rising to 3% at 12 months because of the significant passage of time. Peer effects were guesstimated at 2% at 3 months, rising slightly to 2.5% at 12 months (a more meaningful estimate could be derived from analyzing changes building by building, but that would be much more labor-intensive than this study allows). Possible measurement effects on behavior were thought to be quite small.

Of the four ways in which measurement error could contribute to a perceived increase in recycling behavior, “false negatives” in the “before” monitoring was judged to be the most serious. By definition, we have no actual data on which to base an estimate, but a contribution to the mean of 4% does not seem unreasonable. All the other possible errors of this type were judged to be half or less of that magnitude, with “false negatives” in the “after” monitoring deemed least likely. Combining all of these hypothesized values produces a “3-month” estimated mean of 17.2%, with a standard deviation of 7.8%, and a “12-month” estimated mean of 22.6% with a standard deviation of 8.9%. The outer “tails” of the standard normal curve are defined by multiplying the standard deviation by factors associated with the probability that a sample falling outside a defined range could still be part of the distribution of the “null” as opposed to a statistically distinct outcome. For example, the portion of the normal curve defined by plus or minus the standard deviation multiplied by 1.645 contains 95% of the expected results for that distribution, so there is only a 5% chance that a sample falling outside that range would be identified as a separate outcome when in fact it is part of that distribution. The portion of the normal curve defined by plus or minus the standard deviation multiplied by 1.96 contains 97.5% of the expected results, leaving only a 2.5% of an erroneous conclusion. Further, these chances of error are divided between the two “tails” of the curve, so if one is examining a result at or beyond one end, the chances of error are respectively 2.5% and 1.25%. With this in mind, the high end of our theoretical distribution is 32.5% at 3 months and 39.9% at 12 months. But of course, it is only a theoretical distribution.

### **Calculation of “Null Hypothesis” Parameters from Relevant Groups in Study**

With the theoretical framework above as a backdrop, the next step is to analyze the actual results for all of the study groups except the two key experimental groups (“committers”) to see what estimate of the “null hypothesis” may be derived from them. The eight groups or subgroups to be considered are listed at the top of Attachment 3. We have already concluded that the control group alone, the results from which would have constituted the “null hypothesis” under the original study design, will not suffice. Instead, it is worth exploring the possibility that the eight listed groups are samples of the null hypothesis, capturing various combinations and degrees of the effects or factors discussed earlier.

Several different possible parameters of the 3-month and 12-month distributions that might have produced the eight results we obtained were calculated. First we calculated the means, weighted and unweighted. The unweighted means for all eight

groups combined were 17.4% at 3 months and 23.3% at 12 months. The weighted means were lower, at 13.7% for 3 months and 19.0% for 12 months. Next, we calculated standard deviations and sample standard deviations (a more conservative, thus larger, number which is relevant when sample sizes are fairly small). For the unweighted samples, standard deviations were 7.2% at 3 months and 7.5% at 12 months. Sample standard deviations were 7.7% at 3 months and 8.0% at 12 months. When the samples were weighted by size, standard deviations were 6.0% at 3 months and 6.6% at 12 months.

It seems advisable to adopt a fairly conservative approach to selecting the parameters of our “null hypothesis”, since we have no way of verifying that it has, or even approximates, a normal distribution. Of the various ways of calculating the two key parameters shown in Attachment 3, the most conservative is to use the unweighted (higher) mean and the sample standard deviation. This is illustrated as Approach #2, and produces the conclusion that an experimental result above 39.0% will have only a 1.25% chance of actually being due to the combined factors constituting the null hypothesis. This result is quite close to the conservative theoretical value described in the previous section.

### **Mailing (Brochure) Group Conclusion**

It did not require any sophisticated analysis to determine that the results for the Mailing (or Brochure) treatment group did not demonstrate a meaningful impact for this outreach method. Since this study targeted non-recycling households who would presumably have received one to many mailings previously (since this is a primary outreach methods in most communities), it is logical that one more mailing might have little impact. In fact, the net increase in recycling behavior in the Mailing group was only marginally higher than in the Control group (11.2% versus 10.9% in the “3-month” monitoring, and 13.9% versus 13.2% in the “12-month” monitoring. More significantly, though, as described above there were significantly higher results in all of the various “non-committer” subgroups.

### **Statistical Analysis of Results for “Committer” Subgroups**

The large number of imperfectly defined factors operating in this study made a definitive choice of statistical testing method difficult. As an aid to discussing the results, the graphs in Attachment 4 were developed. They illustrate, first for the 3-month results and then for the 12-month results, the relationship of the observed increase in recycling behavior in the two experimental groups to the background effects which we have synthesized into a “null hypothesis”.

For the 3-month results, the phone group would appear, at 100%, to be well outside any possibility of confusion with the bell curve at left. However, with a tiny sample size (N=4), this result is better described as “strongly suggestive” rather than definitively significant in the statistical sense. The 12-month result for the phone group – clearly NOT statistically significant – reinforces this conclusion. When N is only 4, the

sample standard deviation is larger as a percentage of the mean of the distribution being sampled. In other words, if ANY of the various factors we discussed earlier were to affect one household, the result would shift by 25%. In the unlikely event that two factors shifted two households, the result would potentially change by 50%. In fact, it seems as likely that household turnover or a false positive in the 3-month monitoring are involved in the 12-month drop as it does that there was actually a 75% drop-out rate.

The conclusions are more encouraging for the “door committer” subgroup, for which the sample size (N=30) meets the requirements of most tests of statistical significance. Assuming that the actual distribution of the null hypothesis is near enough to normal that our conservative assumptions account for it, then both the 3-month and the 12-month results seem significant. The 12-month result is somewhat less significant, since it remained the same at 46.7%, while the distribution of our null hypothesis shifted upward.

### **Conclusions**

The limitations of this study, both in original design and events during the course of the study, have made it difficult to conduct any rigorous statistical analysis of the results within the very limited budget and timeframe available. Rather, emphasis has been placed on developing a logical framework for accounting for all the variables in play. With that said, Clear View Consulting concludes that it is most likely that further statistical analysis would conclude that the result for the “door committers” is significant evidence that the outreach technique being tested had a meaningful impact. However, it must be pointed out that the 46.7% measured result cannot reliably be used as an estimate of the effect of the outreach alone, since some of the “background” factors might well have been operating in that subgroup as well. The most likely estimate of the effect of the outreach alone would be the 46.7% less the mean of our null hypothesis, or 29.3% for the 3-month results, and 23.4% for the 12-month results, but it is only an estimate.

Particularly in light of the 12-month results, CVC concludes that the result for the “phone committers” should not be considered statistically significant, but rather strongly suggestive of meaningful results from the outreach effort.

It had been hoped that some statistical analysis of this study’s secondary research question, the drop-out rates among “committers”, would be possible, but the evolution of the study design and the many complicating factors make that even more difficult than addressing the primary research question. The elimination of the originally planned “1-month” monitoring for budgetary reasons made it impossible to distinguish between “committers” who never actually started to recycle, and those who started but dropped out. Statistical analysis of the changes from “3-months” to 12-months” could only be usefully done if the sample sizes had been larger.